

BARBARA DĘBSKA^{*)}, ELWIRA WIANOWSKA^{**)}

Badania właściwości żywicy melaminowej modyfikowanej akryloamidem

Cz. II. PROGNOZOWANIE WŁAŚCIWOŚCI^{***)}

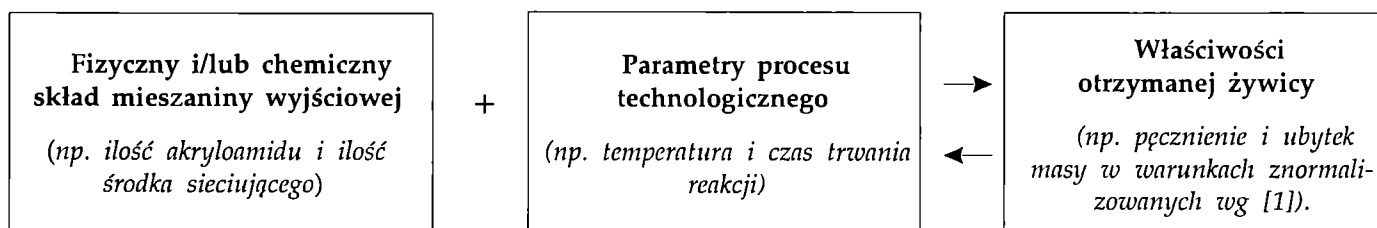
INVESTIGATIONS ON THE PROPERTIES OF AN ACRYLAMIDE-MODIFIED MELAMINE RESIN. PART. II. PREDICTION OF POLYMER PROPERTIES

Summary — Classification and graphical presentation of experimental data sets were used in the SCANKEE system to predict properties of an acrylamide-modified melamine resin. The SCANKEE system is based on a specific formalism of process representation known as the knowledge association (Scheme I). In the planning of experiments this association was used to predict (i) whether an experimental blend of polymers (*viz.*, acrylamide and cross-linking agent amounts) and the contemplated operating parameters (*viz.*, reaction temperature and time) will yield a desirably water-resistant melamine resin (Fig. 2, Table 3) and (ii) what blend composition and what process parameters should be used to prepare a product endowed with assumed properties.

Key words: melamine resins, classification of objects, prediction of properties, water resistance, association of knowledge, analysis of data sets.

Pierwsza część naszej pracy [1] dotyczyła otrzymania żywic melaminowo-acetonowo-formaldehadowych modyfikowanych akryloamidem (żywica MAFA) z roztworów melaminy w reaktywnych rozpuszczalnikach. Głównym celem badań laboratoryjnych była ocena wodoodporności (wyrażonej pośrednio przez pęcz-

zna przedstawić za pomocą ogólnego schematu I. Schemat ten ma szczególne znaczenie w technologii chemicznej lub technologiach pokrewnych np. w metalurgii, przemyśle szkła i ceramiki, przemyśle gumowym, a także w wielu innych dziedzinach nauki, techniki i gospodarki.



Schemat I

nienie i ubytek masy) próbek żywic MAFA. Doświadczalne dane analityczne stały się podstawą komputerowej bazy danych. Wykorzystano w niej specyficzny formalizm reprezentacji wiedzy o procesie technologicznym (nazywany asocjacją wiedzy). Formalizm ten mo-

Każda asocjacja opisuje jedną próbkę żywicy. Zbiór wszystkich asocjacji stanowi bazę danych doświadczalnych. Tabela 1 zawiera fragment zbudowanej bazy danych dotyczący 8 próbek żywicy MAFA (łącznie przebadano 84 różne próbki żywicy). W odniesieniu do każdej próbki dane te określają ilość akryloamidu i stężenie inicjatora rodnikowego — H₂O₂ — w mieszaninie reakcyjnej, warunki utwardzania żywicy (temperaturę i czas) oraz oceniane w tej pracy właściwości próbki żywicy (pęcznienie i ubytek masy).

Zgromadzone w toku prac doświadczalnych dane

^{*)} Politechnika Rzeszowska, Katedra Informatyki Chemicznej, Al. Powstańców Warszawy 6, 35-041 Rzeszów; e-mail: bjdebska@prz.rzeszow.pl.

^{**)} Politechnika Radomska, Katedra Technologii Ogólnej i Syntetycznych Materiałów Obuwniczych, ul. Malczewskiego 22, 26-600 Radom.

^{***)} Cz. I — por. [1].

T a b e l a 1. Fragment źródłowej bazy danych
T a b l e 1. Part of the source database

Numer próbki	Ilość akryloamidu, g	Stężenie inicjatora rodnikowego (H ₂ O ₂) % mas.	Temperatura utleniania, K	Czas utwardzania min	Pęcznienie, %	Ubytek masy, %
1	0,75	0,3	373	20	30,85	12,29
2	0,75	0,3	373	30	30,56	15,71
3	0,75	0,3	373	40	29,85	14,44
4	0,75	0,3	373	50	23,09	9,27
5	0,75	0,3	393	20	25,41	7,10
6	0,75	0,3	393	30	22,67	8,23
7	0,75	0,3	393	40	16,19	7,32
8	0,75	0,3	393	50	19,69	5,28

umożliwiły statystyczną ocenę wyników. Ich interpretacja pozwoliła na sformułowanie następujących wniosków ogólnych (zweryfikowanych i potwierdzonych za pomocą testów statystycznych) [1]:

— czas utwardzania żywicy w istotny sposób wpływa na ubytek masy badanych próbek, czyli na wodoodporność żywicy,

— zależność między czasem utwardzania żywicy, a ubytkiem masy próbek jest odwrotnie proporcjonalna i może być aproksymowana linią prostą, współczynnik korelacji wynosi -0,99.

Stwierdzenie występowania istotnego związku pomiędzy czasem utwardzania a właściwościami żywicy (ubytkiem masy) skłania do poszukiwania innych zależności przyczynowo-skutkowych na podstawie istniejącej już bazy danych (np. związków pomiędzy składem mieszaniny reakcyjnej i temperaturą utwardzania a właściwościami otrzymanego polimeru). Ustalenie takich korelacji pozwala na dobranie optymalnych warunków doświadczalnych i umożliwia prognozowanie właściwości nowych żywic melaminowo-acetonoformaldehydowych. Analiza zależności ukrytych w bazie danych pociąga za sobą konieczność jednoczesnej interpretacji wielu danych doświadczalnych (wielowymiarowych asocjacji wiedzy). Może być ona również prowadzona metodą analizy korelacji (np. metodą kanonicznej analizy korelacyjnej), jednakże w przypadku takich wielowymiarowych zbiorów stosuje się zazwyczaj inne statystyczne metody analizy danych. Zalicza się do nich m.in. analizę składowych głównych, analizę czynnikową, analizę skupień oraz skalowanie wielowymiarowe. W naszych badaniach do oceny bazy danych wykorzystano metody analizy skupień [2].

ANALIZA SKUPIEŃ

Głównym celem analizy skupień jest ustalenie zależności pomiędzy danymi oraz określenie struktury zbioru danych, prowadzące do klasyfikacji wielowymiarowych obiektów (wielowartościowych asocjacji

wiedzy). Klasyfikacja ma na celu wskazanie takich podzbiorów w całym zbiorze danych, które grupują obiekty najbardziej podobne do siebie. Podzbiory te określa się mianem skupień. Podstawowym parametrem w analizie skupień jest odległość pomiędzy obiektami, będąca miarą ich podobieństwa. Istnieje wiele różnych sposobów definiowania odległości i metod tworzenia skupień [2]. Do naszych celów najbardziej użyteczne okazały się odległość euklidesowa i metoda tworzenia skupień zwana SAHN (Sekwencyjna, Aglomeracyjna, Hierarchiczna dająca Niepokrywające się skupienia obiektów) [2, 3]. Metoda została wykorzystana w opracowanym przez nas programie komputerowym. Poza funkcją tworzenia skupień, program ten daje możliwość tworzenia wykresów dwu- i trójwymiarowych, co pozwala na wizualną ocenę wydzielonych klas i znacznie ułatwia interpretację.

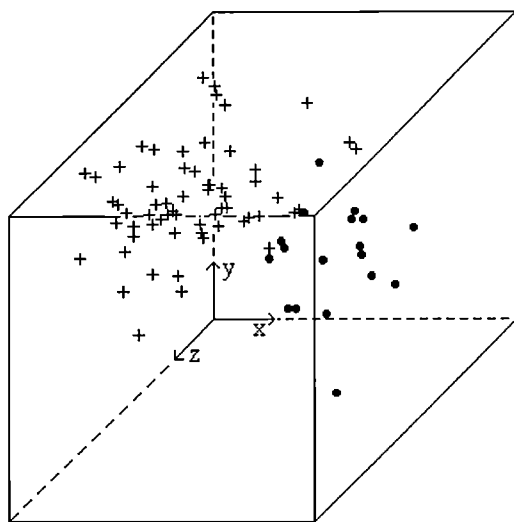
Wszystkie obliczenia związane z analizą skupień wykonaliśmy za pomocą komputera IBM PC (16 MB RAM, Pentium 133 MHz).

PRZETWARZANIE DANYCH

Budową baz danych analitycznych zajmują się zazwyczaj technolodzy z dużym doświadczeniem zawodowym. Inżynier pracujący w zakładzie przemysłowym jest ekspertem w swojej dziedzinie, jednakże często operuje wiedzą, którą trudno mu sformalizować. Dlatego niezbędna jest automatyzacja zarówno w dziedzinie pozyskiwania wiedzy oraz budowania bazy danych, jak i wykrywania nieznanymi wcześniej prawidłowości istniejących między danymi, co umożliwia formułowanie wniosków wynikających z doświadczeń i podejmowanie nowych decyzji.

Zastosowanie odpowiednich programów komputerowych umożliwia automatyzację procedury pobierania i przetwarzania danych. Przykładem takiego narzędzia informatycznego jest system SCANKEE opracowany w Katedrze Informatyki Chemicznej Politechniki Rzeszowskiej [4]. Trzy moduły tego systemu: MKF (*Multistrategy Learning Formalizer*), MLC (*Multistrategy Learning Classifier*) i wykorzystywany przez nas VVT (*Virtual Visualization Tool*) służą do budowy baz wiedzy, przetwarzania zgromadzonych zbiorów i poszukiwania ukrytych regularności pomiędzy danymi.

Uzyskane podczas badań żywic MAFA dane doświadczalne zostały przygotowane w postaci pliku typu ASCII (tabela 1). Zbiór ten poddano procedurze grupowania obiektów, której celem był podział zbioru badanych próbek żywicy na dwie oddzielne klasy. Na tym etapie badań próbowaliśmy wstępnie ustalić czy w badanym zbiorze próbek żywicy można wskazać produkty charakteryzujące się "dobrą" (dużą) i "złą" (małą) wodoodpornością. Próba nienadzorowanej klasyfikacji próbek żywicy (rys. 1) nie doprowadziła do poprawnego podziału zbioru danych na oddzielne klasy, gdyż widoczne dwie grupy obiektów częściowo



Rys. 1. Wynik nienadzorowanej klasyfikacji zbioru próbek żywicy: + — klasa I ("zła"); • — klasa II ("dobra"); x — ubytek masy, %; y — pęcznienie, %; z — stężenie inicjatora rodnikowego (H_2O_2), % mas.

Fig. 1. The result of unsupervised classification of a set of resin samples: + — class I ("bad"), • — class II ("good"), x — weight loss, %, y — swelling, %, z — radical initiator (H_2O_2) concentration, % by wt.

nakładają się na siebie. Rysunek ten, podobnie jak pozostałe, jest schematem poglądowym, a nie wykresem.

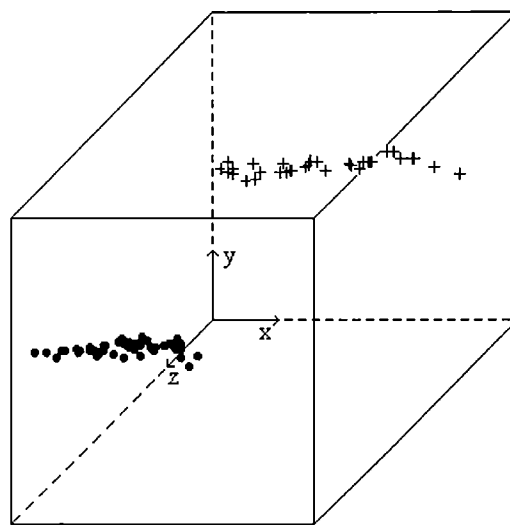
W celu poprawienia wyników grupowania obiektów dokonaliśmy rozszerzenia źródłowej bazy danych (z tabeli 1) o dodatkową kolumnę, w którą wpisana została ocena próbki żywicy. Wartość oceny wynosiła 0 lub 1 i w związku z tym cecha ta posiadała właściwości klasyfikacyjne. Wyznaczenia wartości siódmej cechy dokonano w sposób automatyczny, analizując obliczone wartości ubytku masy badanych próbek. Wygenerowano więc nową bazę danych (tabela 2), w której w odniesieniu do każdej asocjacji wiedzy (każdej próbki polimeru) obliczono wartość dodatkowej, siódmej cechy. W algo-

T a b e l a 2. Fragment bazy danych z tabeli 1 uzupełniony o ocenę próbki

T a b e l e 2. The part of the database of Table 1 and the evaluation of the sample

Numer próbki	Ilość akrylamidu g	Stężenie inicjatora rodnikowego (H_2O_2) % mas.	Temperatura utleniania K	Czas utwardzania min	Pęcznienie %	Ubytek masy %	Ocena próbki
1	0,75	0,3	373	20	30,85	12,29	0
2	0,75	0,3	373	30	30,56	15,71	0
3	0,75	0,3	373	40	29,85	14,44	0
4	0,75	0,3	373	50	23,09	9,27	0
5	0,75	0,3	393	20	25,41	7,10	1
6	0,75	0,3	393	30	22,67	8,23	1
7	0,75	0,3	393	40	16,19	7,32	1
8	0,75	0,3	393	50	19,69	5,28	1

rytmie ustalania wartości tej cechy wykorzystano wartość średnią ubytku masy, ponieważ jest ona najczęściej używana do oceny wyników doświadczenia. Wartość średnia jest liczbą, w przypadku której suma kwadratów odchyień wartości poszczególnych obserwacji jest najmniejsza z możliwych. Obliczona wartość średnia ubytku masy wszystkich próbek wynosiła 8,92%. Podzieliła ona zbiór 84 asocjacji na dwie klasy: w klasie I (wartość cechy siódmej 0) znalazły się te próbki żywicy, których wartość ubytku masy była większa od wartości średniej (co pośrednio świadczyło o "złej" wodoodporności), natomiast w klasie II (wartość cechy siódmej 1) — próbki żywicy charakteryzujące się wartością ubytku masy nie większą niż wartość średnia (tj. "dobrą" wodoodpornością).



Rys. 2. Wizualizacja wyników klasyfikacji próbek żywicy: + — klasa I ("zła"); • — klasa II ("dobra"); x — ubytek masy, %; y — ocena próbki; z — pęcznienie, %

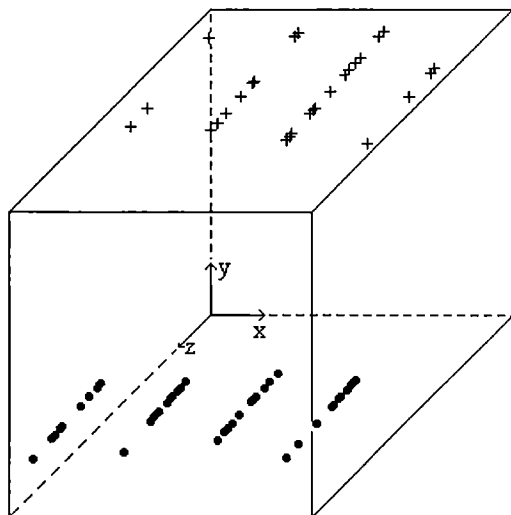
Fig. 2. Visualization of classification of resin samples: + — class I ("bad"), • — class II ("good"), x — weight loss, %, y — evaluation of sample, z — swelling, %

T a b e l a 3. Wynik podziału zbioru próbek żywicy na dwie klasy

Próbki o "złej" wodoodporności (klasa I)	Próbki o "dobrej" wodoodporności (klasa II)
1, 2, 3, 4, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20, 21, 22, 29, 30, 33, 37, 41, 49, 50, 51, 53, 57, 58, 61, 65, 69, 73, 77, 78	5, 6, 7, 8, 15, 16, 23, 24, 25, 26, 27, 28, 31, 32, 34, 35, 36, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, 52, 54, 55, 56, 59, 60, 62, 63, 64, 66, 67, 68, 70, 71, 72, 74, 75, 76, 79, 80, 81, 82, 83, 84

Wyniki klasyfikacji tak przygotowanego zbioru obiektów przedstawiają tabela 3 i rys. 2. Wprowadzenie dodatkowej danej, oceniającej jakość każdej otrzymanej próbki żywicy, doprowadziło do liniowej rozdzielczości zbioru asocjacji. Klasa I ("zła") zawiera 33 obiekty, na-

tomiast klasa II ("dobra") — 51 obiektów. Graficzna prezentacja wyników metodą rzutowania zbioru obiektów na przestrzeń trójwymiarową (rys. 2) potwierdza więc zaproponowany teoretycznie podział zbioru próbek żywicy na dwie klasy. Zmieniając układ i znaczenie osi można zaobserwować dodatkowo podział obiektów wewnątrz klas (rys. 3). Przedstawiona na osi x pierwsza cecha jednoznacznie dzieli zbiór obiektów na cztery podzbiory związane z ilością akryloamidu w żywicy, wynoszącą odpowiednio: 0,25; 0,5; 0,75 lub 1,0 g.



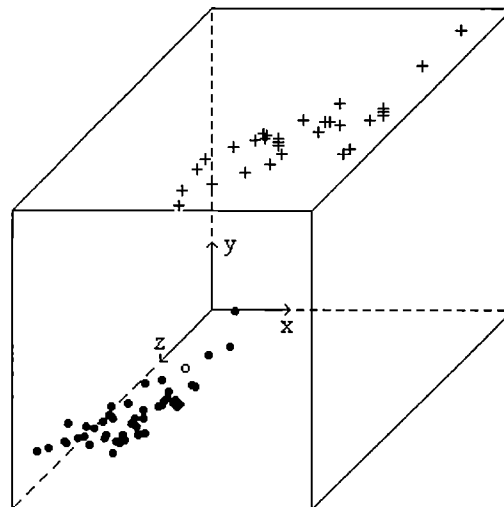
Rys. 3. Graficzna prezentacja podzbiorów wyróżnionych w klasach: + — klasa I ("zła"); • — klasa II ("dobra"); x — ilość akryloamidu, g; y — ocena próbki; z — ubytek masy, %
Fig. 3. Graphical presentation of subsets distinguished in the following classes: + — class I ("bad"), • — class II ("good"), x — amount of acrylamide, g, y — evaluation of sample, z — weight loss, %

Uwzględniając uzyskane wyniki można stwierdzić, że zbiór obiektów, w przypadku których cecha siódma została wyliczona na podstawie średniej wartości ubytku masy, został podzielony na dwa podzbiory charakteryzujące się dużą gęstością skupienia obiektów w każdej z ustalonych grup (rys. 2). Wykorzystana metryka euklidesowa (do wyznaczania odległości pomiędzy obiektami) zapewnia zatem poprawny podział obiektów na dwie rozłączne klasy i odwzorowuje rzeczywistą strukturę zbioru próbek. Dlatego też kolejne opcje programu, tj. "klasyfikacja obiektów" oraz "przewidywanie cech" uruchomiono zakładając, że obliczenia odległości obiektów będzie się wykonywać z zastosowaniem metryki euklidesowej.

Klasyfikacja nieznanego obiektów

Klasyfikacja nieznanego obiektu (w odniesieniu do badań nad metodami otrzymywania żywicy) dotyczy określenia przynależności danej próbki do jednego z dwóch alternatywnych podzbiorów obiektów ce-

chujących się "dobrą" i "złą" wodoodpornością. Klasyfikacja odbywa się metodą KNN (*K-Nearest Neighbours*), na zasadzie określenia przynależności klasowej trzech najbliższych sąsiadów nieznanego obiektu. W celu ich wskazania należy obliczyć odległość pomiędzy obiektem badanym a wszystkimi obrazami należącymi do dwóch klas. Badany obiekt jest przypisywany do tej klasy, do której należy co najmniej dwóch jego najbliższych sąsiadów, albowiem o jego przynależności decyduje większość sąsiadów (stosuje się tutaj metodę głosowania). Wyniki klasyfikacji, przedstawione w postaci graficznej, pozwalają na ocenę, do której klasy należy badana próbka. Oprócz wskazania lokalizacji danego obiektu, moduł graficzny umożliwia również za-



Rys. 4. Lokalizacja przykładowego obiektu z klasy II: + — klasa I ("zła"); • — klasa II ("dobra"); o — obiekt nieznany; x — ubytek masy, %; y — ocena próbki; z — pęcznienie, %
Fig. 4. Localization of an illustrative class II object: + — class I ("bad"), • — class II ("good"), o — unknown object, x — weight loss, %, y — evaluation of sample, z — swelling, %

T a b e l a 4. Klasyfikacja nieznanego obiektu metodą KNN
T a b l e 4. Classification of an unknown object by the KNN method

Próbka	Ilość akryloamidu, g	Stężenie inicjatora rodnikowego (H ₂ O ₂) % mas.	Temperatura K	Czas utwardzania żywicy min	Pęcznienie, %	Ubytek masy %	Ocena próbki polimeru
Nieznanymy obiekt	1,0	0,07	393	30	24,37	8,43	1
Punkt sąsiadujący 66	1,0	0,07	373	30	23,12	8,33	1
Punkt sąsiadujący 67	1,0	0,07	373	40	24,23	7,33	1
Punkt sąsiadujący 75	1,0	0,10	393	20	24,89	8,04	1

poznanie się z wartościami cech opisującymi jego trzech najbliższych sąsiadów (rys. 4, tabela 4).

Ponieważ baza danych zawierała wyniki uzyskane doświadczalnie, przeprowadziliśmy dodatkowe badania wpływu dokładności wyników na poprawność klasyfikacji. Poprawność działania procedury klasyfikacji obiektów sprawdziliśmy w odniesieniu do dwóch rodzajów danych (określonych z różną precyzją):

— zbiory zawierające takie próbki, w przypadku których dane doświadczalne definiujące klasyfikowaną asocjację były wyznaczone z niewielkim błędem (obliczone odchylenie standardowe ubytków masy próbek żywicy miało wartość mniejszą niż 0,5),

— zbiory zawierające takie próbki, w przypadku których dane doświadczalne definiujące klasyfikowaną asocjację były obciążone dużym błędem (obliczone odchylenie standardowe ubytków masy próbek żywicy miało wartość większą niż 0,9).

W obydwu przypadkach zbiory testujące poprawność prognozowania zawierały po 10 przykładów. W odniesieniu do pierwszego zbioru, wartości przewidywanych klas obiektów zgadzały się w 100% z klasą rzeczywistą. Natomiast w drugim przypadku przynależność klasowa jedynie siedmiu badanych obiektów została ustalona poprawnie.

PODSUMOWANIE

Przykładem wykorzystania omówionych metod pozyskiwania wiedzy z baz danych analitycznych było ich zastosowanie do prognozowania niektórych parametrów produkcji nowych wyrobów. W interesującym nas zagadnieniu, główną korzyścią wynikającą z analizy skupień zbioru próbek żywicy była możliwość nauce komputera prognozowania właściwości żywicy MAFA. Wartości danej cechy obiektu (ubytku masy

próbki) przewidywano metodą najbliższego sąsiada — KNN. Obliczona wartość średniej ważonej była przedstawiana jako prognozowana cecha danego obiektu. Zaprogramowany (w systemie SCANKEE) algorytm klasyfikacyjny SAHN wykorzystywano do rozwiązania dwóch głównych problemów występujących podczas planowania doświadczenia chemicznego, tzn. problemu klasyfikacji żywic otrzymanych z danej mieszaniny substratów i poddanej działaniu określonych parametrów procesu technologicznego oraz problemu prognozowania, w wyniku którego następuje taki dobór optymalnego składu mieszaniny substratów i parametrów procesu, aby żywica miała założone właściwości.

Wykonane przez nas testy procedur klasyfikacji i prognozowania potwierdzają istotny wpływ danych (obciążonych dużym błędem) na wyniki wnioskowania statystycznego.

W trzeciej (ostatniej) części pracy podejmiemy próbę wyjaśnienia mechanizmu otrzymywania żywic melaminowo-acetonowo-formaldehadowych na podstawie zarejestrowanych widm NMR badanych próbek.

LITERATURA

1. Dębska B., Malinowska E.: *Polimery* 1999, **44**, 608.
2. Marek T.: "Analiza skupień w badaniach empirycznych. Metody SAHN", PWN, Warszawa 1989.
3. Dębska B.: "The application of visual methods and conversions of analytical data in the systems of multcategory object's classification", Proceedings of the Workshop "Intelligent Information Systems", Dęblin, 2—5 czerwca 1996, str. 127—135.
4. Hippe Z.: *Computers & Chemistry* 1998, **22**, nr 1, 133.

Otrzymano 29 III 1999 r.